The PAVOQUE corpus as a resource for analysis and synthesis of expressive speech

Ingmar Steiner[1–3], Marc Schröder[3], Annette Klepp[1,3]
steiner@coli.uni-saarland.de

[1]Multimodal Computing & Interaction

[2]Saarland University

[3]German Research Center for Artificial Intelligence

Overview

We announce the release of the PAVOQUE corpus, a single-speaker, multi-style database of German speech, designed for analysis and synthesis of expressive speech. The corpus has been previously used for voice conversion [5] and expressive text-to-speech synthesis [1, 4].

The *full* corpus data is now being made available to the public, under a Creative Commons license. It is hosted at

`https://github.com/marytts/pavoque-data`

Speaker and recordings

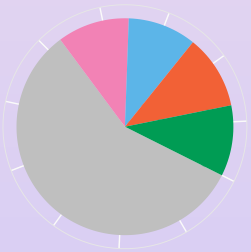
src: www.stefan-roettig.de

Stefan Röttig, a male native speaker of German trained as a professional actor and baritone opera singer, was hired to produce the corpus.

The recordings were carried out in a sound-proof studio, over multiple sessions, with a sampling rate of 44.1 kHz at 24 bit per sample.

All utterances were automatically transcribed using MaryTTS [3]; the phone-level segmentation was manually verified by phonetically trained research assistants.

Corpus composition

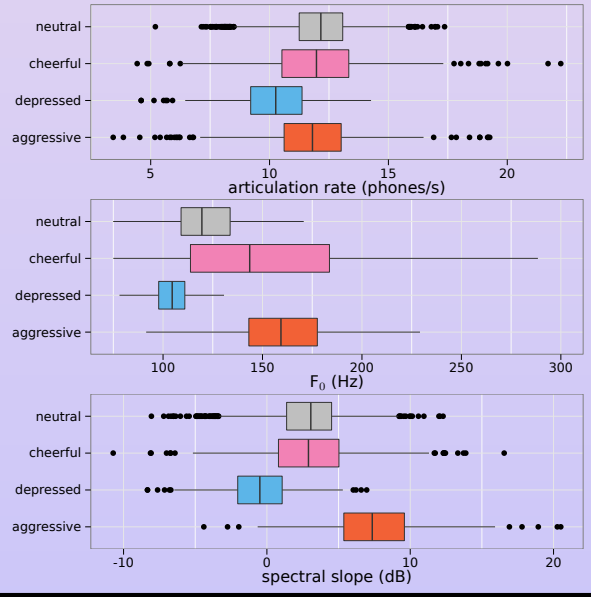


The bulk of the corpus consists of general-domain sentences (A) automatically extracted from Wikipedia using a greedy algorithm optimizing for phonetic and prosodic coverage [2]; these were spoken in a neutral, "news-reading" style. 375 more of these (B) are common to all styles. A number of domain-specific utterances (C) were spoken as well.

set	neutral	cheerful	depressed	aggressive	poker
A	2639	25	25	25	25
B	375	375	375	375	375
C	112	184	156	201	175
total	3126	584	556	601	575
time	321 min	46 min	55 min	45 min	49 min

Overall, 5442 utterances (8 h 37 min) are available in five different speaking styles.

Selected statistics



References

[1] P. Gebhard, M. Schröder, M. Charfuelan, C. Endres, M. Kipp, S. Pammi, M. Rumpler, and O. Türk. "IDEAS4Games: building expressive virtual characters for computer games". In: *8th International Conference on Intelligent Virtual Agents (IVA)*. Tokyo, Japan, 2008, pp. 426–440. DOI: 10.1007/978-3-540-85483-8_43.

[2] A. Hunecke. "Optimal Design of a Speech Database for Unit Selection Synthesis". Diploma thesis. Saarbrücken, Germany: Saarland University, 2007.

[3] M. Schröder, M. Charfuelan, S. Pammi, and I. Steiner. "Open source voice creation toolkit for the MARY TTS platform". In: *Interspeech*. Florence, Italy, 2011, pp. 3253–3256.

[4] I. Steiner, M. Schröder, M. Charfuelan, and A. Klepp. "Symbolic vs. acoustics-based style control for expressive unit selection". In: *7th ISCA Tutorial and Research Workshop on Speech Synthesis (SSW)*. Kyoto, Japan, 2010, pp. 114–119.

[5] O. Türk and M. Schröder. "Evaluation of expressive speech synthesis with voice conversion and copy resynthesis techniques". In: *IEEE Transactions on Audio, Speech, and Language Processing* 18.5 (2010), pp. 965–973. DOI: 10.1109/TASL.2010.2041113.